《Python数据分析与应用》教学大纲

**课程名称：**Python数据分析与应用

**课程类别：**必修

**适用专业：**大数据技术类相关专业

**先修课程：**Python编程基础

**总学时：**64学时（其中理论28学时，实验36学时）

**总学分：**4.0学分

# 课程的性质

大数据时代已经到来，在商业、经济及其他领域中基于数据和分析去发现问题并做出科学、客观的决策越来越重要。数据分析技术将帮助企业用户在合理时间内获取、管理、处理以及整理海量数据，为企业经营决策提供积极的帮助。数据分析作为一门前沿技术，广泛应用于物联网、云计算、移动互联网等战略新兴产业。有实践经验的数据分析人才已经成为了各企业争夺的热门。为了推动我国大数据，云计算，人工智能行业的发展，满足日益增长的数据分析人才需求，特开设Python数据分析与应用课程。

# 课程的基本要求

理论上，要求学生掌握数据分析的概念、流程；了解数据分析的应用场景和工具；掌握数据分析常用库的作用；掌握数值计算、统计分析、数据预处理、数据可视化、模型构建的常用方法。

技能上，要求学生掌握使用Python进行数值计算、统计分析、数据预处理、数据可视化、模型构建的操作方法，能将知识点揉入竞赛网站、企业所得税、餐饮企业等真实示例中，实现学以致用；并能基于TipDM大数据挖掘建模平台，使用拖拽式、流程化、去编程化的新技术，解决实际案例。

思政上，将习近平新时代中国特色社会主义思想、数据安全法等思政要素融汇到教学中，引发学生对新技术、新形势的思考，促进经济、行业的进一步发展；培养学生的科学探索、独立思考、全局意识和思辨能力，以及职业素养、信息保护意识等。

# 教学条件

Python 3.8.5+ Anaconda3 2020.11+ scikit-learn 0.22

# 课程学时分配

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 教学内容 | 理论学时 | 实验学时 | 其他 |
| 1 | 第1章 Python数据分析概述 | 2 | 1 |  |
| 2 | 第2章 NumPy数值计算基础 | 2 | 2 |  |
| 3 | 第3章 pandas统计分析基础 | 2 | 2 |  |
| 4 | 第4章 使用pandas进行数据预处理 | 3 | 4 |  |
| 5 | 第5章 Matplotlib、seaborn、pyecharts数据可视化基础 | 3 | 4 |  |
| 6 | 第6章 使用sklearn构建模型 | 6 | 4 |  |
| 7 | 第7章 竞赛网站用户行为分析 | 3 | 5 |  |
| 8 | 第8章 企业所得税预测分析 | 3 | 5 |  |
| 9 | 第9章 餐饮企业客户流失预测 | 3 | 5 |  |
| 10 | 第10章 基于TipDM大数据挖掘建模平台实现客户流失预测 | 1 | 4 |  |
|  | 总计 | 28 | 36 |  |

# 教学内容及学时安排

### 理论教学

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **章节名称** | **主要内容** | **教学目标** | **学时** |
| 1 | Python数据分析概述 | 1. 掌握数据分析的概念 2. 掌握数据分析的流程 3. 了解数据分析的应用场景 4. 了解数据分析的常用工具 5. 了解Python数据分析的优势 6. 了解Python数据分析的常用类库 7. 了解Python的Anaconda发行版 8. 在Windows操作系统上安装Anaconda 9. 在Linux系统上安装Anaconda 10. 掌握Jupyter Notebook的基础功能 11. 掌握Jupyter Notebook的高级功能 | 1. 掌握数据分析的概念、流程与应用场景 2. 了解Python常用的数据分析库 3. 掌握Windows / Linux系统下Anaconda安装 4. 掌握Jupyter Notebook的常用功能 | 2 |
| 2 | NumPy数值计算基础 | 1. 创建数组对象 2. 生成随机数 3. 通过索引访问数组 4. 变换数组的形态 5. 创建NumPy矩阵 6. 掌握ufunc函数 7. 读写文件 8. 使用数组进行简单的统计分析 | 1. 掌握NumPy创建多维数组与生成随机数的方法 2. 掌握数组的索引与变换 3. 掌握NumPy中数组矩阵的运算及通用函数的基本使用方法 4. 掌握NumPy读写文件的方法和常用的统计分析的函数 | 2 |
| 3 | pandas统计分析基础 | 1. 读写文本文件 2. 读写Excel文件 3. 读写数据库数据 4. 查看DataFrame的常用属性 5. 查改增删DataFrame数据 6. 描述分析DataFrame数据 7. 转换字符串时间为标准时间 8. 提取时间序列数据信息 9. 加减时间数据 10. 使用groupby方法拆分数据 11. 使用agg方法聚合数据 12. 使用apply方法聚合数据 13. 使用transform方法聚合数据 14. 使用povit\_table函数创建透视表 15. 使用crosstab函数创建交叉表 | 1. 掌握常见的数据读取方式 2. 掌握DataFrame常用属性与方法 3. 掌握基础时间数据处理方法 4. 掌握分组聚合的原理与方法 5. 掌握透视表与交叉表的制作 | 2 |
| 4 | 使用pandas进行数据预处理 | 1. 堆叠合并数据 2. 主键合并数据 3. 重叠合并数据 4. 检测与处理重复值 5. 检测与处理缺失值 6. 检测与处理异常值 7. 离差标准化数据 8. 标准差标准化数据 9. 小数定标标准化数据 10. 哑变量处理类别型数据 11. 离散化连续型数据 | 1. 掌握数据合并的原理与方法 2. 掌握数据清洗的基本方法 3. 掌握基本数据标准化的方法 4. 掌握常用的数据转换方法 | 3 |
| 5 | Matplotlib、seaborn、pyecharts数据可视化基础 | 1. 掌握pyplot的基础语法 2. 设置pyplot的动态rc参数 3. 使用Matplotlib绘制进阶图形绘制折线图 4. 熟悉seaborn绘图基础 5. 使用seaborn绘制基础图形 6. 熟悉pyecharts绘图基础 7. 使用pyecharts绘制交互式图形 | 1. 掌握pyplot常用的绘图参数的调节方法 2. 掌握seaborn绘制基础图形的方法 3. 掌握seaborn绘制基础图形的方法 4. 掌握pyecharts绘制交互式图形的方法 | 3 |
| 6 | 使用scikit-learn构建模型 | 1. 加载datasets模块中的数据集 2. 将数据集划分为训练集和测试集 3. 使用sklearn转换器进行数据预处理与降维 4. 使用sklearn估计器构建聚类模型 5. 评价聚类模型 6. 使用sklearn估计器构建分类模型 7. 评价分类模型 8. 使用sklearn估计器构建回归模型 9. 评价回归模型 | 1. 掌握sklearn转换器的使用方法 2. 掌握sklearn估计器的使用方法 3. 掌握聚类模型的构建与评价 4. 掌握分类模型的构建与评价 5. 掌握回归模型的构建与评价 | 6 |
| 7 | 竞赛网站用户行为分析 | 1. 了解竞赛网站背景 2. 认识用户行为分析 3. 熟悉竞赛网站用户行为分析的步骤与流程 4. 特征值转换 5. 用户识别 6. 数据清洗 7. 网页分类 8. 构造特征 9. 构造特征 10. 了解K-Means聚类算法 11. 使用K-Means聚类算法进行用户分群 | 1. 熟竞赛网站用户行为分析的步骤和流程 2. 预处理竞赛网站用户访问数据 3. 掌握K-Means算法的基本原理与使用方法 4. 对竞赛网站用户进行分群 | 3 |
| 8 | 企业所得税预测分析 | 1. 分析企业所得税预测背景 2. 了解企业所得税预测的方法 3. 熟悉企业所得税预测的步骤与流程 4. 了解相关性分析 5. 计算Pearson相关系数 6. 了解Lasso回归方法 7. 选取关键特征 8. 了解灰色预测算法 9. 了解SVR算法 10. 预测企业所得税 | 1. 熟悉企业所得税预测的步骤和流程 2. 掌握相关性分析方法与应用 3. 掌握使用Lasso模型选取特征的方法 4. 掌握灰色预测的原理与应用 5. 掌握支持向量回归算法的基本原理与应用 | 3 |
| 9 | 餐饮企业客户流失预测 | 1. 了解餐饮企业客户流失预测背景 2. 认识餐饮客户流失预测 3. 熟悉餐饮企业客户流失预测的步骤与流程 4. 数据探索 5. 查看重复值 6. 处理异常值 7. 处理异常值 8. 构建客户流失特征了解决策树算法 9. 了解支持向量机算法 10. 预测餐饮企业客户流失 | 1. 熟悉餐饮企业客户流失预测的步骤与流程 2. 了解决策树算法的基本原理与应用 3. 了解支持向量机算法的基本原理与应用 | 3 |
| 10 | 基于TipDM数据挖掘建模平台实现客户流失预测 | 1. 了解平台的界面、访问方式和特点 2. 了解【实训库】模块的功能 3. 了解【数据连接】模块的功能 4. 了解【实训数据】模块的功能 5. 了解【我的实训】模块的功能 6. 了解【系统算法】模块的功能 7. 了解【个人算法】模块的功能 8. 掌握平台配置客户流失预测案例的流程和步骤 9. 数据源配置 10. 数据预处理 11. 构建模型 | 1. 了解平台的相关概念、特点和功能 2. 掌握平台配置客户流失预测案例的流程和步骤 | 1 |
| 学 时 合 计 | | | | 28 |

### 实验教学

|  |  |  |  |
| --- | --- | --- | --- |
| **序号** | **实验项目名称** | **实验要求** | **学时** |
| 1 | Python数据分析环境搭建 | 1. 在Windows/Linux系统上安装Anaconda 2. 掌握Jupyter Notebook的常用功能 | 1 |
| 2 | NumPy数值计算基础 | 1. 创建NumPy数组对象ndarray 2. 查看ndarray的常用属性 3. 花式索引ndarray 4. 变换ndarray的形态 5. 创建NumPy矩阵并使用；使用常见ufunc 6. 使用NumPy读写文件 | 2 |
| 3 | pandas统计分析基础 | 1. 读写文本文件 2. 读写Excel文件 3. 读写数据库数据 4. 查看DataFrame的常用属性 5. 查改增删DataFrame数据 6. 描述分析DataFrame数据 7. 转换字符串时间为标准时间 8. 提取时间序列数据信息 9. 加减时间数据 10. 使用groupby方法拆分数据 11. 使用agg，apply，transform方法聚合数据 12. 制作透视表 13. 制作交叉表 | 2 |
| 4 | 使用pandas进行数据预处理 | 1. 堆叠、主键、重叠合并数据 2. 检测与处理重复值，缺失值，异常值 3. 离差标准化、标准差标准化 4. 小数定标标准化数据；哑变量处理类别型数据 5. 离散化连续型数据 | 4 |
| 5 | Matplotlib、seaborn、pyecharts数据可视化基础 | 1. 掌握pyplot的基本绘图语法 2. 设置pyplot的动态rc参数 3. 绘制散点图 4. 绘制折线图 5. 绘制柱形图 6. 绘制饼图 7. 绘制箱线图 8. 熟悉seaborn绘图基础 9. 绘制热力图 10. 绘制分类散点图 11. 绘制线性回归拟合图 12. 熟悉pyecharts绘图基础 13. 绘制3D散点图 14. 绘制漏斗图 15. 绘制词云图 | 4 |
| 6 | 使用sklearn构建模型 | 1. 加载datasets模块自带数据集 2. 划分数据集 3. 使用sklearn转换器进行数据预处理与降维 4. 构建与评价聚类模型 5. 构建与评价分类模型 6. 构建与评价回归模型 | 4 |
| 7 | 竞赛网站用户行为分析 | 1. 预处理竞赛网站用户访问数据 2. 使用K-Means聚类算法进行用户分群 3. 评价K-Means聚类模型 | 5 |
| 8 | 企业所得税预测分析 | 1. 分析企业所得税数据特征的相关性 2. 使用Lasso回归选取企业所得税数据特征的关键特征 3. 使用灰色预测和SVR构建企业所得税预测模型 4. 评价SVR模型 | 5 |
| 9 | 餐饮企业客户流失预测 | 1. 探索数据 2. 查看与处理餐饮企业数据中的重复值、异常值、缺失值 3. 构建客户流失特征 4. 使用决策树和支持向量机进行餐饮客户流失预测 | 5 |
|  | 基于TipDM大数据挖掘建模平台实现客户流失预测 | 1. 配置客户流失预测案例的数据源 2. 基于TipDM数据挖掘建模平台对餐饮企业数据进行预处理 3. 基于TipDM数据挖掘建模平台构建餐饮客户流失预测模型 | 4 |
| 学 时 合 计 | | | 36 |

# 考核方式

突出学生解决实际问题的能力，加强过程性考核。课程考核的成绩构成 = 平时作业（10%）+ 课堂参与（20%）+ 期末考核（70%），期末考试建议采用开卷形式，试题应包括基本概念、绘图、分组聚合、数据合并、数据清洗、数据变换、模型构建等部分，题型可采用判断题、选择、简答、应用题等方式。

# 教材与参考资料

### 教材

曾文权，张良均．Python数据分析与应用（第2版）（微课版）[M]．北京：人民邮电出版社．2022．

### 参考资料

[1] 黄红梅，张良均．Python数据分析与应用[M]．北京：人民邮电出版社．2018．

[2] 张良均，谭立云．Python数据分析与挖掘实战（第2版）[M]．北京：机械工业出版社．2019．

[3] 张健，张良均．Python编程基础[M]．北京：人民邮电出版社．2018．